arXiv:1011.1328v1 [math.ST] 5 Nov 2010

General model selection estimation of a periodic regression with a Gaussian noise

Victor Konev* Serguei Pergamenchtchikov †

Abstract

This paper considers the problem of estimating a periodic function in a continuous time regression model with an additive stationary gaussian noise having unknown correlation function. A general model selection procedure on the basis of arbitrary projective estimates, which does not need the knowledge of the noise correlation function, is proposed. A non-asymptotic upper bound for $\mathcal{L}_2$-risk (oracle inequality) has been derived under mild conditions on the noise. For the Ornstein-Uhlenbeck noise the risk upper bound is shown to be uniform in the nuisance parameter. In the case of gaussian white noise the constructed procedure has some advantages as compared with the procedure based on the least squares estimates (LSE). The asymptotic minimaxity of the estimates has been proved. The proposed model selection scheme is extended also to the estimation problem based on the discrete data applicably to the situation when high frequency sampling can not be provided.

Key words: model selection procedure, periodic regression, oracle inequality, non-parametric regression, improved estimation

*Department of Applied Mathematics and Cybernetics, Tomsk State University, Lenin street 36, 634050 Tomsk, Russia. e-mail: vvkonev@mail.tsu.ru

†Laboratoire de Mathématiques Raphaël Salem, UMR 6085 CNRS, Université de Rouen, Avenue de l'Université, BP.12, 76801 Saint Etienne du Rouvray (France). email: Serge.Pergamenchtchikov@univ-rouen.fr

1 Introduction

Consider a regression model in continuous time

$$\mathrm{d}y_t \,=\, S(t)\mathrm{d}t \,+\, \mathrm{d}\xi_t\,, \tag{1}$$

where $S(t)$ is an unknown 1-periodic function in the space $\mathcal{L}_2[0,1]$, $(\xi_t)_{t\ge 0}$ is a continuous gaussian process with zero mean and such that for each $n \ge 1$ the stochastic integral $\int_0^n f(t)\mathrm{d}\xi_t$ is well-defined for any non-random function f from $\mathcal{L}_2[0,n]$. The correlation function of noise ξ_t is unknown.

This process can be modeled in different ways.

Example 1. *ξ_t is a scalar non-explosive Ornstein-Uhlenbeck process defined by the equation*

$$\mathrm{d}\xi_t = \theta\xi_t\mathrm{d}t + \mathrm{d}w_t\,, \tag{2}$$

where $(w_t)_{t\ge 0}$ is a standard brownian motion and $\theta \le 0$ is unknown parameter; the initial value $\xi_0 \sim \mathcal{N}(0, 1/2|\theta|)$ if $\theta < 0$ and $\xi_0 = 0$ if $\theta = 0$.

Example 2. *ξ_t is a stationary autoregressive process of order $q \ge 2$ satisfying the stochastic differential equation*

$$\xi_t^{(q)} = \theta_1\xi_t^{(q-1)} + \ldots + \theta_q\xi_t + \dot{w}_t\,. \tag{3}$$

Here $(\dot{w}_t)_{t\ge 0}$ is a white gaussian noise and the unknown vector $\theta = (\theta_1, \ldots, \theta_q)'$ belongs to stability region of the process

$$\mathcal{A} = \{\theta \in \mathbb{R}^q \,:\, \max_{1\le i\le q} Re\lambda_i(\theta) < 0\}\,, \tag{4}$$

where $(\lambda_i(\theta))_{1\le i\le q}$ are eigenvalues of the matrix

$$A = A(\theta) = \begin{pmatrix} \theta_1 & \ldots & \theta_q \\ & I_{q-1} & 0 \end{pmatrix}\,; \tag{5}$$

I_q is the identity matrix of order q.

Models of type (1) and their discrete-time analogues have been studied by a number of authors (see, Efroimovich (1999), Liptser and Shyraev (1974), Konev and Pergamenshchikov (2003), Nemirovskii (2000) and references therein). The estimation problem of periodic signal $S(t)$ in model (1)–(2) has been thoroughly studied in the case, when $(\xi_t)_{t\ge 0}$ is a white gaussian noise (see, for example, Ibragimov and Hasminskii (1981) for details and further references).

A discrete-time counterpart of model (1)–(2) was applied in the econometrical problems for modeling the consumption as a function of income Golfeld and Quandt (1972).

As is well known, the problem of nonparametric estimation of $S(t)$ comprises the following three statements: the function estimation at a fixed point t_0, estimation in the uniform metric and in the integral metric. The first two problems are usually solved by making use of the kernel and local polynomial estimates. This paper focuses on the third setting with the quadratic metric. The estimation in the integral metric is based, as a rule, on the projective estimates which were first proposed in Chenstov (1962) for estimating the distribution density in a scheme of i.i.d. observations. The heart of this method is to approximate the unknown function with a finite Fourier series. Applying the projective estimates to the regression model (1) with a white noise leads to the optimal convergence rate in $\mathbf{L}_2(0,1)$ provided that the smoothness of S is known (see for example Ibragimov and Hasminskii (1981)). Another adaptive approach based on the model selection method (see for example, Barron et al. (1999), Baraud (2000), Birgé and Massart (2001) and Fourdrinier and Pergamenshchikov (2007)) enables one to study this problem in the nonasymptotic setting when the smoothness of function S is unknown. It should be noted that this method can be used also for model (1) under the condition that the correlation function $\mathbf{E}\xi_t\xi_s$ is exactly known and besides the unknown function S belongs to the subspace spanned by its eigenfunctions (see, Theorem 1, p. 11 in Birgé and Massart (2001)). In our case, when the noise correlation function is unknown, this method can not be applied. This paper develops a general model selection method for the regression scheme (1) with unknown correlation properties.

Note that the usual nonasymptotic selection model procedure proposed in Barron et al. (1999), Baraud (2000), Birgé and Massart (2001) is based on the least square estimators (LSE) which, as was shown in Goloubev (1982) and Pinsker (1981), are not efficient in the problem of nonparametric regression. Our approach is close to the general model selection method proposed in Fourdrinier and Pergamenshchikov (2007) for discrete time models with spherically symmetric errors which allows one to use any projective estimators in the model selection procedure including the LSE. In Section 2 we propose a general model selection procedure for a regression scheme in continuous time (1) with unknown correlation structure of the gaussian noise. In Theorem 1, under some loose conditions on the noise, we establish a nonasymptotic upper bound for the quadratic risk in which the principal term is minimal over the set of all admissible basic estimates. The inequalities of this type are usually called *oracle*.

In the case of the Ornstein-Uhlenbeck noise (2), the risk upper bound is shown to be uniform in the nuisance parameter (Corollary 2).

The rest of the paper is organized as follows. In Section 3 we consider case of white gaussian noise ξ_t and show that the possibility to choose different projective estimators in the procedure may lead to a sharper upper bound for the mean square estimation accuracy. In Section 4 the upper bound and the lower bounds for the minimax quadratic risk are obtained under the assumption that the smoothness of S is unknown. In Section 6 we consider the estimation problem for the regression model (1) assuming that it is accessible for observations only at discrete times $t_k = k/p$, $k = 0, 1, \ldots$. Such observation scheme is more appropriate in a number of applications, where one can not provide high frequency data sampling. Theorems 5 establishs the nonasymptotic oracle inequalities in this case. Appendix contains some technical results.

2 Nonasymptotic estimation

In this section we consider the estimation problem for the model (1) in nonasymptotic setting, i.e. assuming that the estimator of S is based on the observations $(y_t)_{0\le t\le n}$ with a fixed duration n. For this we apply the general model selection approach proposed in Fourdrinier and Pergamenshchikov (2007) for the discrete-time regression model.

First we introduce some notations. Let $\mathcal{X}$ be the Hilbert space of square integrable 1-periodic functions on $\mathbb{R}$ with the usual scalar product

$$(x, y) \, = \, \int_0^1 x(t)\, y(t)\, \mathrm{d}t$$

and $(\phi_j)_{j\ge 1}$ be a system of orthonormal functions in $\mathcal{X}$, i.e. $(\phi_i, \phi_j) = 0$, if $i \ne j$ and $\|\phi_i\|^2 = (\phi_i, \phi_i) = 1$.

Then we impose the following additional conditions on the noise $(\xi_t)_{t\ge 0}$ in (1). Assume that

$\mathbf{C}_1$) *For each $n \ge 1$ and $k \ge 1$ the vector $\zeta(n) = (\zeta_1(n), \ldots, \zeta_k(n))'$ with components*

$$\zeta_j(n) \, = \, \frac{1}{\sqrt{n}} \int_0^n \phi_j(t)\, \mathrm{d}\xi_t \tag{6}$$

is gaussian with non-degenerate covariance matrix $B_{k,n} = \mathbf{E}\, \zeta(n)\zeta'(n)$.

$\mathbf{C}_2$) *The maximal eigenvalues of matrices $B_{k,n}$ satisfy the following inequality*

$$\sup_{k\ge 1}\sup_{n\ge 1}\lambda_{\max}(B_{k,n})\;\le\;\lambda^*\,,$$

where λ^* is some known positive constant.

Processes (2) and (3) in Examples 1–2, as is shown in Lemmas 2–3, satisfy condition $\mathbf{C}_1$). Condition $\mathbf{C}_2$) is satisfied for process (2) with $\lambda^*=2$. Condition $\mathbf{C}_2$) holds also for process (3) provided that the value of vector θ belongs to the following compact set

$$K_\delta=\left\{\theta\in\mathcal{A}\,:\,\max_{1\le i\le q}\,\mathrm{Re}\,\lambda_i(\theta)\le-\delta\,,\quad|A(\theta)|\le\delta^{-1}\right\}\,,\tag{7}$$

where $0<\delta<1$ is a known constant; $|\cdot|$ stands for the euclidean norm of matrix. Under this assumption process (3) satisfies condition $\mathbf{C}_2$) with

$$\lambda^*=\lambda^*(\delta)=\frac{2}{\delta^2}F^*(\delta)\,J^*(\delta)\,,\tag{8}$$

where

$$F^*(\delta)=\frac{q}{2\delta}+\frac{2q}{\delta^3}\sum_{j=1}^{q-1}\frac{(2j)!}{(j!)^2\delta^{4j}}\quad\text{and}\quad J^*(\delta)=\frac{1}{\delta}+\frac{2}{\delta^2}\sum_{j=1}^{q-1}\frac{2^j}{\delta^{2j}}\,.$$

Let $\mathbb{N}$ be the set of positive integer numbers, i.e. $\mathbb{N}=\{1,2,\ldots\}$. Denote by $\mathcal{M}$ some finite set of finite subsets of $\mathbb{N}$ and by $(\mathcal{D}_m)_{m\in\mathcal{M}}$ a family of linear subspaces of $\mathcal{X}$ such that

$$\mathcal{D}_m=\{x\in\mathcal{X}\,:\,x=\sum_{j\in m}\lambda_j\phi_j\,,\lambda_j\in\mathbb{R}\}\,.$$

Let $d_m=\dim\mathcal{D}_m$ be the number of elements in a subset m. Denote by S_m the projection of S on $\mathcal{D}_m$, i.e.

$$S_m=\sum_{j\in m}\alpha_j\phi_j\,,\quad\alpha_j=(S,\phi_j)\,.\tag{9}$$

To estimate the function S in (1) we will apply a general model selection approach. It requires first to choose some class of projective estimators $\widetilde{S}_m$ of S_m, which may be any measurable functions of observations $(y_t)_{0\le t\le n}$ taking on values in $\mathcal{D}_m$. For example, one can take the LSE $\widehat{S}_m$ of S, which is the minimizer, with respect to $x\in\mathcal{D}_m$, of the quantity

$$\gamma_n(x)=\|x\|^2\,-\,2\,\frac{1}{n}\int_0^n x(t)\,\mathrm{d}y_t\tag{10}$$

and has the form

$$\widehat{S}_m = \sum_{j\in m} \widehat{\alpha}_j\, \phi_j\,, \quad \widehat{\alpha}_j \;=\; \frac{1}{n}\int_0^n \phi_j(t)\,\mathrm{d}y_t\,. \tag{11}$$

Let $(l_m)_{m\in\mathcal{M}}$ be a sequence of prior weights such that $l_m \geq 1$ for all $m \in \mathcal{M}$. We set

$$l^* = \sum_{m\in\mathcal{M}} e^{-l_m\, d_m}\,. \tag{12}$$

Further one needs a penalty term on the set $\mathcal{M}$. We take it in the form suggested in Birgé and Massart (2001). We define the penalty term as

$$P_n(m) \;=\; \rho\,\frac{l_m\, d_m}{n} \quad \text{with} \quad \rho = \,4\lambda^*\,\frac{z_*^2}{z_*-1}\,, \tag{13}$$

where z_* is the maximal root of the equation $\ln z = z-2$ which is approximately equal to $z_* \approx 3,1462$.

Minimizing the penalized empirical contrast $\gamma_n(\widetilde{S}_m)+P_n(m)$ with respect to $m\in\mathcal{M}$ one finds

$$\widetilde{m} = \mathrm{argmin}_{m\in\mathcal{M}}\{\gamma_n(\widetilde{S}_m)+P_n(m)\} \tag{14}$$

and obtains the model selection procedure $\widetilde{S}_{\widetilde{m}}$ corresponding to a specific class of projective estimators $(\widetilde{S}_m)_{m\in\mathcal{M}}$. For the LSE family $(\widehat{S}_m)_{m\in\mathcal{M}}$, this yields $\widehat{S}_{\widehat{m}}$ with

$$\widehat{m} = \mathrm{argmin}_{m\in\mathcal{M}}\{\gamma_n(\widehat{S}_m)+P_n(m)\}\,. \tag{15}$$

Our first result is the following.

Theorem 1. *Assume that the conditions* $\mathbf{C}_1)$–$\mathbf{C}_2)$ *are fulfilled for the noise in* (1)*. Then for any class of projective estimators* $(\widetilde{S}_m)_{m\in\mathcal{M}}$ *the general model selection procedure* $\widetilde{S}_{\widetilde{m}}$ *satisfies the following oracle inequality*

$$\mathbf{E}_S\,\|\widetilde{S}_{\widetilde{m}}-S\|^2 \;\leq\; \inf_{m\in\mathcal{M}}\,\widetilde{a}_m(S)\;+\frac{\lambda^*\tau_0}{n}\,, \tag{16}$$

where $\mathbf{E}_S$ *denotes the expectation with respect to the distribution of* (1) *given* S *,*

$$\widetilde{a}_m(S) = 3\mathbf{E}_S\|\widetilde{S}_m-S\|^2 + 16\lambda^* z_*\,\frac{d_m\, l_m}{n} \quad \textit{and} \quad \tau_0 = \frac{16 l^* z_*}{z_*-1}\,.$$

The proof of Theorem 1 is given in the Appendix.

Remark 1. *It will be noted that the choice of the coefficient ρ in the penalty term* (13)*, as will be shown in the proof of the theorem, provides the minimal value of the principal term $\widetilde{a}_m(S)$.*

Now we will find the upper bound (16) for the LSE model selection procedure $\widehat{S}_{\widehat{m}}$ defined by (11) and (15). To this end we have to calculate the accuracy of $\widehat{S}_m$ for any $m \in \mathcal{M}$. We have

$$\mathbf{E}_S \, \|\widehat{S}_m - S\|^2 = \|S_m - S\|^2 + \mathbf{E}_S \, \|\widehat{S}_m - S_m\|^2$$
$$= \|S_m - S\|^2 + \sum_{j \in m} \mathbf{E}_S \, (\widehat{\alpha}_j - \alpha_j)^2 \,,$$

where S_m is given in (9). Moreover, the condition $\mathbf{C}_2$) yields

$$\mathbf{E}_S \, (\widehat{\alpha}_j \, - \alpha_j)^2 = \frac{1}{n^2} \, \mathbf{E}_S \, \left(\int_0^n \, \phi_j(t) \, \mathrm{d}\xi_t \right)^2 \leq \frac{\lambda^*}{n} \,.$$

Therefore

$$\mathbf{E}_S \, \|\widehat{S}_m - S\|^2 \, \leq \, \|S_m - S\|^2 + \lambda^* \frac{d_m}{n} \,.$$

Thus, we obtain the following result.

Corollary 1. *Under the conditions $\mathbf{C}_1$) and $\mathbf{C}_2$) the LSE model selection procedure $\widehat{S}_{\widehat{m}}$, defined by* (11) *and* (15)*, satisfies the inequality*

$$\mathbf{E}_S \, \|\widehat{S}_{\widehat{m}} - S\|^2 \, \leq \, \inf_{m \in \mathcal{M}} \, \widehat{a}_m(S) \, + \, \frac{\lambda^* \tau_0}{n} \,, \tag{17}$$

where

$$\widehat{a}_m(S) = 3\|S_m - S\|^2 + \tau_1 \lambda^* \, \frac{d_m l_m}{n} \,, \quad \tau_1 = 3 + 16 z_* \,.$$

Consider the upper bound in (17) in more detail for the model (1)-(2).

Corollary 2. *For the model* (1)–(2) *the LSE model selection procedure $\widehat{S}_{\widehat{m}}$, defined by* (11) *and* (15) *with $\lambda^* = 2$ satisfies, for any $\theta \leq 0$, the inequality*

$$\mathbf{E}_S \, \|\widehat{S}_{\widehat{m}} - S\|^2 \, \leq \, \inf_{m \in \mathcal{M}} \, \widehat{b}_m(S) \, + \, \frac{2\tau_0}{n} \,, \tag{18}$$

where τ_0 is given in (16)*,*

$$\widehat{b}_m(S) = 3\|S_m - S\|^2 + 2\tau_1 \, \frac{d_m \, l_m}{n} \,.$$

Remark 2. *It will be noted that for the model* (1)–(2) *the LSE model selection procedure satisfies the oracle inequality uniformly in the nuisance parameter θ including the boundary of the stationarity region of the Ornstein-Uhlenbeck process, i.e. $\theta = 0$.*

3 The improvement of LSE.

In this section we consider a special case of the model (1)–(2) when $\theta = 0$, i.e.

$$\mathrm{d}y_t \,=\, S(t)\,\mathrm{d}t \,+\, \mathrm{d}w_t\,. \tag{19}$$

By applying the improvement method proposed in Fourdrinier and Pergamenshchikov (2007) we will show that the upper bound in the oracle inequality can be lessened by a proper choice of the projective estimators. Let us introduce a class of estimators of the form

$$S^*_m(t) = \widehat{S}_m(t) + \Psi_m(\widehat{S}_m)(t)\,. \tag{20}$$

Here Ψ_m is a function from $\mathbb{R}^d$ into $\mathcal{D}_m$, i.e.

$$\Psi_m(x)(t) = \sum_{j\in m} v_j(x)\,\phi_j(t)\,, \quad x\in\mathbb{R}^d \tag{21}$$

and $(v_j(\cdot))_{j\in m}$ are $\mathbb{R}^d\to\mathbb{R}$ functions such that $\mathbf{E}_S\, v_j^2(\widehat{\alpha})<\infty$, where $\widehat{\alpha} = (\widehat{\alpha}_j)_{j\in m}$ is the vector with the components $\widehat{\alpha}_j$ defined in (11). The functions $v(x) = (v_j(x))_{j\in m}$ will be specified below. Let

$$\Delta_m(S) \,=\, \mathbf{E}_S\,\|S^*_m - S_m\|^2 \,-\, \mathbf{E}_S\,\|\widehat{S}_m - S_m\|^2\,. \tag{22}$$

It is easy to check that

$$\Delta_m(S) \,=\, 2\,\mathbf{E}_S\,(\Psi_m\,,\,\widehat{S}_m - S_m) \,+\, \mathbf{E}_S\,\|\Psi_m\|^2\,. \tag{23}$$

This function can be found explicitly for the model (19).

Lemma 1. *Let S^*_m be defined by (20)–(21) with continuously differentiable functions v_j such that $\mathbf{E}_S\, v_j^2(\widehat{\alpha})<\infty$. Then $\Delta_m(S) \,=\, \mathbf{E}_S\, L(\widehat{\alpha})$, where*

$$L(x) \,=\, \frac{2}{n}\,\mathrm{div}\,v(x) \,+\, \|v(x)\|^2\,. \tag{24}$$

Proof. From (11), (20), one has

$$\widehat{\alpha}_j \,=\, \frac{1}{n}\int_0^n \phi_j(t)\,\mathrm{d}y_t \,=\, \alpha_j \,+\, \frac{1}{n}\int_0^n \phi_j(t)\,\mathrm{d}w_t\,,$$

where $\alpha_j = \int_0^1 \phi_j(t)S(t)\,\mathrm{d}t$. Therefore the vector $\widehat{\alpha} = (\widehat{\alpha}_j)_{j\in m}$ has a normal distribution $\mathcal{N}(\alpha, n^{-1}I_d)$, where $\alpha = (\alpha_j)_{j\in m}$ and I_d is the unit matrix of

order d. This enables one to find the explicit expression for the first term in the right-hand side of (23). Indeed,

$$J = \mathbf{E}_S\,(\Psi_m, \widehat{S}_m - S_m) = \mathbf{E}_S \sum_{j\in m} v_j(\widehat{\alpha})\,(\widehat{\alpha}_j - \alpha_j)$$
$$= \mathbf{E}_S\,(v(\widehat{\alpha})\,,\,\widehat{\alpha}-\alpha) = \int_{\mathbb{R}^d} (v(u)\,,\,u-\alpha)\,g(\|u-\alpha\|^2)\,\mathrm{d}u\,,$$

where

$$g(a) = \left(\frac{n}{2\pi}\right)^{d/2} e^{-na/2}\,. \tag{25}$$

Making the spherical changes of the variables yields

$$J = \int_0^\infty \int_{\mathcal{S}_{r,d}} (v(u)\,,\,u-\alpha)\,\nu_{r,d}(\mathrm{d}u)\,g(r^2)\,\mathrm{d}r$$
$$= \int_0^\infty \int_{\mathcal{S}_{r,d}} (v(u)\,,\,e(u))\,\nu_{r,d}(\mathrm{d}u)\,r\,g(r^2)\,\mathrm{d}r\,,$$

where $\mathcal{S}_{r,d} = \{u \in \mathbb{R}^d \,:\, \|u-\alpha\| = r\}$, $\nu_{r,d}(\cdot)$ is the superficial measure on the sphere $\mathcal{S}_{r,d}$ and $e(u) = (u-\alpha)/\|u-\alpha\|$ is a normal vector to this sphere. By applying the Ostrogradsky–Stokes divergence theorem we obtain that

$$J = \int_0^\infty \int_{\mathcal{B}_{r,d}} \operatorname{div} v(u)\,\mathrm{d}u\,r\,g(r^2)\,\mathrm{d}r$$

with $\mathcal{B}_{r,d} = \{u \in \mathbb{R}^d \,:\, \|u-\alpha\| \le r\}$. By the Fubini theorem and the definition of g in (25) one gets

$$J = \frac{1}{2}\int_{\mathbb{R}^d} \int_{\|u-\alpha\|^2}^\infty g(a)\,\mathrm{d}a \operatorname{div} v(u)\,\mathrm{d}u$$
$$= \frac{1}{n}\int_{\mathbb{R}^d} g(\|u-\alpha\|^2)\,\operatorname{div} v(u)\,\mathrm{d}u = \frac{1}{n}\,\mathbf{E}_S \operatorname{div} v(\widehat{\alpha})\,.$$

This leads to the assertion of Lemma 1. $\square$

In particular for $d_m > 2$, if one takes

$$v(u) = -\frac{(d_m-2)u}{n\|u\|^2}\,, \quad \text{then} \quad L(u) = -\frac{(d_m-2)^2}{n^2\|u\|^2}\,,$$

and hence, $\Delta_m(S) < 0$, that is, the estimate (20) outperforms the least squares esimate (11) in the approximation of S_m. This allows to improve the model selection procedure by making use of the estimates (20) instead of the least squares $\{\widehat{S}_m\}$. As a direct consequence of Theorem 1, one obtains the following result for the improved model selection procedure $S^*_{m^*}$.

Theorem 2. *For the model* (19) *the improvement model selection procedure* $S^*_{m^*}$ *defined by (14) with* $\widetilde{S}_m = S^*_m$ *and* $\lambda^* = 2$ *satisfies the inequality*

$$\mathbf{E}_S \, \|S^*_{m^*} - S\|^2 \leq \inf_{m\in\mathcal{M}} u^*_m(S) + \frac{2\tau_0}{n} \,, \tag{26}$$

where τ_0 *is given in* (16),

$$u^*_m(S) \, = \, 3\,\mathbf{E}_S \, \|S^*_m - S\|^2 \, + \, 32\, z_* \, \frac{d_m \, l_m}{n} \,.$$

4 Asymptotic estimation

4.1 The risk upper bound

In this section we consider the asymptotic estimation problem for the model (1). To this end, we additionally assume that all functions in the orthonormal system $(\phi_j)_{j\geq 1}$ are 1-periodic and the unknown function $S(\cdot)$ in the model (1) belongs to the following functional class

$$\Theta_{\beta,r} \, = \, \{S \in \mathcal{C}(\mathbb{R}) \cap \mathcal{X} \, : \, \max_{n\geq 1} n^{2\beta} \, \varsigma_n(S) \, \leq r^2\} \,. \tag{27}$$

Here $\mathcal{C}(\mathbb{R})$ denotes the set of all continuous $\mathbb{R} \to \mathbb{R}$ functions and

$$\varsigma_n(S) = \sum_{j=n}^{\infty} s^2_j \,, \tag{28}$$

where $(s_j)_{j\geq 1}$ are the Fourier coefficients for the basis $(\phi_j)_{j\geq 1}$, i.e.

$$s_j = (S, \phi_j) = \int_0^1 S(t) \, \phi_j(t) \, \mathrm{d}t \,;$$

$\beta > 0$ and $r > 0$ are unknown constants.

Similarly to Galtchouk and Pergamenshchikov (2006) we define now the risk for an estimator $\widetilde{S}_n$ (a measurable function of the observation $(y_t)_{0\leq t\leq n}$ in (6)) as follows

$$\mathcal{R}_n(\widetilde{S}_n, \beta) \, = \, \sup_{S\in\Theta_{\beta,r}} \, \sup_{Q\in\mathcal{P}_\kappa} \mathbf{E}_{S,Q} \|\omega_n(\widetilde{S}_n - S)\|^2 \,, \tag{29}$$

where $\omega_n = \omega_n(\beta) = n^{\frac{\beta}{2\beta+1}}$. Here $\mathcal{P}_\kappa$ is some class of distributions Q (in the space $\mathbf{C}[0, +\infty)$) of the noise process $(\xi_t)_{t\geq 0}$ satisfying conditions $\mathbf{C}_1$)

and $\mathbf{C}_2$) with $\lambda^* = \lambda^*(Q) \le \kappa < \infty$ for some known fixed parameter κ. In addition, this class is assumed to include the Wiener distribution Q_0. The second index in $\mathbf{E}_{S,Q}$ denotes that the expectation is taken with respect to the distribution of the process (1) corresponding to the noise distribution Q.

Note that for the model (1)–(2) $\mathcal{P}_\kappa$ is the class of distributions of the processes (2) with $\theta \le 0$. In this case $\kappa = 2$. For the model (1)–(3)

$$\mathcal{P}_\kappa = \mathcal{Q}_\delta \cup \{Q_0\}\,,$$

where $\mathcal{Q}_\delta$ is the family of distributions of the processes (3) of order $q \ge 2$ with the parameters belonging to the set (7) for some $0 < \delta < 1$. In this case $\kappa = \max(2, \lambda^*(\delta))$, where $\lambda^*(\delta))$ is given in (8).

We will apply the ordered variable model selection procedure (see, Barron et al. (1999), p. 315), for which $\mathcal{M} = \{m_1, \dots, m_n\}$ with $m_i = \{1, \dots, i\}$, therefore $d_{m_i} = i$. Then

$$\mathcal{D}_m = \{x \in \mathcal{X}_n \,:\, x = \sum_{j=1}^{i} \alpha_j\, \phi_j\,, \alpha_j \in \mathbb{R}\}\,.$$

For the ordered variable model selection procedure one can take $l_m = 1$ for all $m \ge 1$ and find

$$l^* = \sum_{i=1}^{n} e^{-d_{m_i}} \le \frac{1}{e-1}\,.$$

In the sequel we denote by $\widehat{S}^\kappa_{\widehat{m}}$ the LSE model selection procedure (11), (15) replacing λ^* by κ. Now we will show that the risk (29) for this procedure is finite.

Theorem 3. *The estimator $\widehat{S}^\kappa_{\widehat{m}}$ satisfies the following asymptotic inequality*

$$\limsup_{n\to\infty} \sup_{\beta>0} \mathcal{R}_n(\widehat{S}^\kappa_{\widehat{m}}, \beta) < \infty\,. \tag{30}$$

Proof. Taking into account (17) one gets, for any $Q \in \mathcal{P}_\kappa$,

$$\begin{aligned}
\mathbf{E}_{S,Q}\|\widehat{S}^\kappa_{\widehat{m}} - S\|^2 &\le \inf_{m\in\mathcal{M}_n} \left(3\,\|S_m - S\|^2 \,+\, \tau_1\,\kappa\frac{d_m}{n}\right) + \frac{\kappa\tau_0}{n} \\
&\le \inf_{1\le i\le n} \left(3\,\|S_{m_i} - S\|^2 \,+\, \tau_1\,\kappa\,\frac{i}{n}\right) + \frac{\kappa\tau_0}{n}\,.
\end{aligned}$$

Further, for any function S from $\Theta_{\beta,r}$, one has

$$\|S_{m_i} - S\|^2 = \sum_{j=i+1}^{\infty} s_j^2 \,\le\, r^2\, i^{-2\beta}\,.$$

Therefore for each $1 \le i \le n$

$$\sup_{Q\in\mathcal{P}_\kappa} \mathbf{E}_{S,Q}\|\widehat{S}^{\kappa}_{\widehat{m}} - S\|^2 \,\le\, \left(3\, r^2\, i^{-2\beta} \,+\, \tau_1 \kappa \frac{i}{n}\right) + \frac{\tau_0 \kappa}{n}\,.$$

Substituting $i = i_n = [n^{\frac{1}{2\beta+1}}] + 1$ leads to (30). $\square$

4.2 The risk lower bound

Now we study the lower bound for the risk (29). We assume that the orthogonal system $(\phi_j)_{j\ge1}$ in (27) is trigonometric, i.e.

$$\phi_1(x) \,\equiv\, 1\,, \quad \text{and for} \quad j \ge 2 \quad \phi_j(x) \,=\, \sqrt{2}\,\mathrm{Tr}_j(2\pi[j/2]x)\,, \qquad (31)$$

where $\mathrm{Tr}_j(x) = \cos x$ for even $j \ge 1$ and $\mathrm{Tr}_j(x) = \sin x$ for odd $j \ge 1$; $[a]$ is the integer part of a.

Theorem 4. *The lower bound of the risk* (29) *over all estimates is strictly positive, i.e. for any $\beta \ge 1$*

$$\underline{\lim}_{n\to\infty} \inf_{\widetilde{S}_n} \mathcal{R}_n(\widetilde{S}_n, \beta) > 0\,. \qquad (32)$$

Proof. In order to show (32) it suffices to check this inequality for the model (19), i.e. that for any $\beta \ge 1$

$$\underline{\lim}_{n\to\infty} \inf_{\widetilde{S}_n} \sup_{S\in\Theta_{\beta,r}} \mathbf{E}_{S,Q_0}\|\omega_n(\widetilde{S}_n - S)\|^2 \,>\, 0\,. \qquad (33)$$

To this end we apply the method proposed in Fourdrinier and Pergamenshchikov (2007) to our case. First, we construct an auxiliary parametric class of functions in the set $\Theta_{\beta,r}$. Let $\beta = k + \alpha$ with $k = [\beta]$ and $0 \le \alpha < 1$. Let $V(\cdot)$ be $k+1$ times continuously differentiable function such that $V(u) = 0$ for $|u| \ge 1$ and $\int_{-1}^{1} V^2(u)\,\mathrm{d}u = 1$. Let $m = [n^{\frac{1}{2\beta+1}}]$ and Γ_δ be a cube in $\mathbb{R}^m$ of the form

$$\Gamma_\delta \,=\, \{z = (z_1, \ldots, z_m)' \in \mathbb{R}^m \,:\, |z_i| \le \delta\,, \quad 1 \le j \le m\}\,,$$

where $\delta = \nu/\omega_n$, $\nu > 0$. Now, viewing the function $V(\cdot)$ as a kernel, one introduces a parametric class of 1-periodic functions $(S_z)_{z\in\Gamma_\delta}$ where

$$S_z(t) = \sum_{j=1}^{m} z_j \psi_j(t)\,, \quad 0 \le t \le 1; \tag{34}$$

$(\psi_j(t))_{1\le j\le m}$ are 1-periodic functions defined on the interval $[0,1]$ as $\psi_j(t) = V\left(\frac{t-a_j}{h}\right)$ with $h = 1/2m$ and $a_j = (2j-1)/2m$.

It will be observed that, for $0 \le i \le k-1$ and $z \in \Gamma_\delta$,

$$\sup_{0\le t\le 1} |S_z^{(i)}(t)| \le 2^i \sup_{|a|\le 1} |V^{(i)}(a)| \, \frac{\nu}{n^{(\beta-i)/(2\beta+1)}} \to 0 \quad \text{as} \quad n\to\infty\,.$$

In order to check the second condition in (62), we estimate the increment of kth derivative of $S_z(\cdot)$. For any $0 \le s\,,\, t \le 1$ and $z \in \Gamma_\delta$, one has

$$|S_z^{(k)}(t) \,-\, S_z^{(k)}(s)| \le \frac{\nu}{\omega_n\, h^k}\, \Delta_m\,, \tag{35}$$

where

$$\Delta_m \,=\, \sum_{j=1}^{m} \left| V^{(k)}\left(\frac{t-a_j}{h}\right) \,-\, V^{(k)}\left(\frac{s-a_j}{h}\right)\right|\,.$$

If s and t belong to the same interval, that is, $a_{j_0} - h \le s \,\le\, t \le a_{j_0} + h$, then putting $V^* = \sup_{|a|\le 1} |V^{(k+1)}(a)|$ one obtains

$$\begin{aligned}\Delta_m \,&=\, \left| V^{(k)}\left(\frac{t-a_{j_0}}{h}\right) \,-\, V^{(k)}\left(\frac{s-a_{j_0}}{h}\right)\right| \\ &\le\, 2\,V^*\,\frac{|t-s|}{2h} \,\le\, 2^{1-\alpha}\,V^*\,\frac{|t-s|^\alpha}{h^\alpha}\end{aligned} \tag{36}$$

for each $0 \le \alpha \le 1$. If s and t belong to different intervals, that is,

$$a_{j_0} - h \le s \le a_{j_0} + h \,\le\, a_{j_1} - h \le t \le a_{j_1} + h\,, \quad j_0 < j_1\,,$$

then setting $s^* = a_{j_0} + h$ and $t_* = a_{j_1} - h$, similarly to (36), one gets

$$\begin{aligned}\Delta_m \,&=\, \left| V^{(k)}\left(\frac{t-a_{j_1}}{h}\right) \,-\, V^{(k)}\left(\frac{t_*-a_{j_1}}{h}\right)\right| \\ &+ \left| V^{(k)}\left(\frac{s-a_{j_0}}{h}\right) - V^{(k)}\left(\frac{s^*-a_{j_0}}{h}\right)\right| \\ &\le\, 2^{1-\alpha}\,V^*\,\frac{1}{h^\alpha}\,\left(|t-t_*|^\alpha \,+\, |s-s^*|^\alpha\right) \,\le\, 2^{2-\alpha}\,\frac{1}{h^\alpha}\,V^*\,|t-s|^\alpha\,.\end{aligned}$$

From here and (35)–(36) we come to the estimate

$$|S_z^{(k)}(t) \,-\, S_z^{(k)}(s)| \,\le\, 2^{2+k}\,\nu\,V^*\,\frac{m^\beta}{\omega_n}\,|t-s|^\alpha\,.$$

Therefore (see Lemma 5 in Appendix 6.4) there exist $\nu > 0$ and $n_0 \ge 1$ such that $S_z \in \Theta_{\beta,r}$ for all $z \in \Gamma_\delta$ and $n \ge n_0$. Further, we introduce the prior distribution on Γ_δ with the density

$$\pi(z) = \pi(z_1, \ldots, z_m) = \prod_{l=1}^m \pi_l(z_l)\,, \quad \pi_l(u) = \frac{1}{\delta} G\left(\frac{u}{\delta}\right)\,.$$

The function $G(u) = G_* e^{-\frac{1}{1-u^2}}$ for $|u| \le 1$ and $G(u) = 0$ for $|u| \ge 1$, where G_* is a positive constant such that $\int_{-1}^1 G(u)\,\mathrm{d}u = 1$.

Let $\widetilde{S}_n(\cdot)$ be an estimate of $S(\cdot)$ based on observations $(y_t)_{0\le t\le n}$ in (1). Then for any $n \ge n_0$ we can estimate with below the supremum in (33) as

$$\sup_{S\in\Theta_{\beta,r}} \mathbf{E}_{S,Q_0} \|\widetilde{S}_n - S\|^2 > \int_{\Gamma_\delta} \mathbf{E}_{S_z,Q_0} \|\widetilde{S}_n - S_z\|^2\,\pi(z)\,\mathrm{d}z\,.$$

Moreover, by the definition of S_z, we obtain

$$\|\widetilde{S}_n - S_z\|^2 \ge \sum_{l=1}^m (\widetilde{z}_l - z_l)^2 \int_0^1 \psi_l^2(t)\,\mathrm{d}t = h \sum_{l=1}^m (\widetilde{z}_l - z_l)^2\,,$$

where $\widetilde{z}_l = \int_0^1 \widetilde{S}_n(x)\,\psi_l(x)\,\mathrm{d}x/\|\psi_l\|^2$. Therefore,

$$\sup_{S\in\Theta_{\beta,r}} \mathbf{E}_{S,Q_0}\,\|\widetilde{S}_n - S\|^2 \ge h \sum_{l=1}^m \Lambda_l\,, \tag{37}$$

where $\Lambda_l = \int_{\Gamma_\delta} \mathbf{E}_{S_z} (\widetilde{z}_l - z_l)^2 \pi(z)\mathrm{d}z$. To apply now lemma 6 we note that in this case

$$\zeta_l(z) = \int_0^n \psi_l(t)\,\mathrm{d}y_t \,-\, \int_0^n S_z(t)\,\psi_l(t)\mathrm{d}t$$

and, therefore, $A_l = \mathbf{E}_{S_z,Q_0}\,\zeta_l^2(z) = \int_0^n \psi_l^2(t)\,\mathrm{d}t = nh$. Moreover, in this case

$$B_l = \int_{-\delta}^{\delta} \frac{(\dot{\pi}_l(u))^2}{\pi_l(u)}\mathrm{d}u = \delta^{-2} I_G \quad ; \quad I_G = 8 \int_0^1 u^2 (1-u^2)^{-4} G(u)\,\mathrm{d}u\,.$$

Thus, by the inequality (66), one obtains that

$$\sup_{S\in\Theta_{\beta,r}} \mathbf{E}_{S,Q_0}\|\widetilde{S}_n - S\|^2 \ge \frac{1}{2m}\sum_{l=1}^{m}\frac{1}{nh+\omega_n^2\nu^{-2}I_G}$$
$$= \frac{1}{2nh+2\omega_n^2\nu^{-2}I_G}\,.$$

This immediately implies (33). Hence Theorem 4. $\square$

5 Estimation based on discrete data.

The model selection procedure developed in Section 2 is intended for continuous time observations. However, in a number of applied problems high frequency sampling can not be provided. In this section, we consider the estimation problem for model (1) on the basis of observations $(y_{t_j})_{0\le j\le np}$ of the process $(y_t)_{t\ge 0}$ at discrete times $t_j = j/p$, where p is a given odd number. To solve this problem, we will modify the model selection procedure of Section 2. Let $\mathcal{X}_p$ be the set of all 1-periodic functions $x:\mathbb{R}\to\mathbb{R}$ with the scalar product

$$(x,z)_p \;=\; \frac{1}{p}\sum_{j=1}^{p} x(t_j)\,z(t_j)\,,\quad x,z\in\mathcal{X}_p\,. \tag{38}$$

Let $(\phi_j)_{1\le j\le p}$ be an orthonormal basis in $\mathcal{X}_p$, i.e. $(\phi_i,\phi_j)_p=0$, if $i\ne j$ and $\|\phi_i\|_p^2=1$. One can use, for example, the trigonometric basis (31).

Assume that the noise $(\xi_t)_{t\ge 0}$ in (1) is such that

$\mathbf{C}_1^*$) *The vector $\zeta^*(n)=(\zeta_1^*(n),\ldots,\zeta_p^*(n))'$ with components*

$$\zeta_l^*(n) \;=\; \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\phi_l(t_j)\,\Delta\xi_{t_j}\,,\quad \Delta\xi_{t_j}=\xi_{t_j}-\xi_{t_{j-1}}\,, \tag{39}$$

is gaussian with non-degenerate covariance matrix $B_{n,p}^*=\mathbf{E}\,\zeta^*(n)(\zeta^*(n))'$;

$\mathbf{C}_2^*$) *The maximal eigenvalues of matrices $B_{n,p}^*$ are uniformly bounded :*

$$\sup_{n\ge1}\sup_{p\ge1}\lambda_{\max}(B_{n,p}^*)\;\le\;\lambda^*\,,$$

where λ^* is some known positive constant.

Conditions $\mathbf{C}_1^*$), $\mathbf{C}_2^*$) are satisfied for processes (2), (3) (cf. Lemmas 2–50).

Now we denote by $\mathcal{M}_p$ some set of subsets of $\{1,\ldots,p\}$ and by $(\mathcal{D}_{m,p})_{m\in\mathcal{M}_p}$ a family of linear subspaces of $\mathcal{X}_p$ such that

$$\mathcal{D}_{m,p} \,=\, \{x\in\mathcal{X}_p \,:\, x=\sum_{j\in m}\lambda_j\phi_j\,,\lambda_j\in\mathbb{R}\}\,.$$

Let $S_{m,p}$ denote the projection of S on $\mathcal{D}_{m,p}$ in $\mathcal{X}_p$ and $\widetilde{S}_{m,p}$ denote an estimator of $S_{m,p}$, i.e. a measurable function of the observations $(y_{t_j})_{0\le j\le np}$ taking on values in $\mathcal{D}_{m,p}$. One can use, for example, the LSE $\widehat{S}_{m,p}$ for $S_{m,p}$, which is defined as the minimizer with respect to $x\in\mathcal{D}_{m,p}$ of the distance

$$\frac{1}{np}\sum_{k=1}^{np}\left(\frac{\Delta y_{t_k}}{\Delta t_k}\,-\,x(t_k)\right)^2$$

that is , the quantity

$$\gamma_{n,p}(x)=\|x\|_p^2\,-\,2\,\frac{1}{n}\sum_{k=1}^{np}x(t_k)\,\Delta y_{t_k} \tag{40}$$

and has the form

$$\widehat{S}_{m,p}=\sum_{j\in m}\widehat{\alpha}_{j,p}\,\phi_j\,,\quad \widehat{\alpha}_{j,p}\,=\,\frac{1}{n}\sum_{k=1}^{np}\phi_j(t_k)\,\Delta y_{t_k}\,. \tag{41}$$

Let the penalty term $P_n(m)$ be defined, as before, by (13). Then the model selection procedure, corresponding to a family of projective estimators $(\widetilde{S}_{m,p})_{m\in\mathcal{M}_p}$, is defined as $\widetilde{S}_{\widetilde{m}_p,p}$ where

$$\widetilde{m}_p=\mathrm{argmin}_{m\in\mathcal{M}_p}\{\gamma_{n,p}(\widetilde{S}_{m,p})+P_n(m)\}\,. \tag{42}$$

In the case of the LSE family $(\widehat{S}_{m,p})_{m\in\mathcal{M}_p}$, it will be $\widehat{S}_{\widehat{m}_p,p}$.

As a measure of accuracy of the approximation of a 1-periodic function S of continuous argument t by its values on the $(t_j)_{1\le j\le p}$, we will use the function

$$H_p(S)\,=\,\frac{1}{p}\sum_{l=1}^{p}h_l^2(S)\,,\quad h_l(S)\,=\,\frac{1}{\Delta t_l}\int_{t_{l-1}}^{t_l}\left(S(t)-S(t_l)\right)\mathrm{d}t\,. \tag{43}$$

The following theorem gives the oracle inequality for a general model selection procedure $\widetilde{S}_{\widetilde{m}_p,p}$ based on the discrete time observations.

Theorem 5. *Assume that the conditions* $\mathbf{C}_1^*$)–$\mathbf{C}_2^*$) *hold. Then the estimator* $\widetilde{S}_{\widetilde{m}_p,p}$ *satisfies the oracle inequality*

$$\mathbf{E}_S\,\|\widetilde{S}_{\widetilde{m}_p,p}-S\|_p^2 \le \inf_{m\in\mathcal{M}_p}\,\widetilde{a}_{m,p}(S)+8H_p(S)+\frac{2\tau_0\lambda^*}{n}\,, \tag{44}$$

where

$$\widetilde{a}_{m,p}(S)\;=\;7\mathbf{E}_S\,\|\widetilde{S}_{m,p}-S\|_p^2\;+\;32\lambda^* z_*\,\frac{l_m d_m}{n}\,.$$

Now we obtain the oracle inequality (44) for the least square model selection procedure $\widehat{S}_{\widehat{m}_p,p}$. To this end, we have to calculate the estimation accuracy of $\widehat{S}_{m,p}$ for $S_{m,p}$, which is the projection of S on $\mathcal{D}_{m,p}$, i.e.

$$S_{m,p}=\sum_{j\in m}\alpha_{j,p}\,\phi_j\,,\quad \alpha_{j,p}=(S,\phi_j)_p=\frac{1}{p}\sum_{k=1}^{p}S(t_k)\,\phi_j(t_k)\,.$$

First of all, we note that in this case

$$\widehat{\alpha}_{j,p}\;-\;\alpha_{j,p}\;=\;\frac{1}{p}\sum_{k=1}^{p}\phi_j(t_k)\,h_k(S)\;+\;\frac{1}{n}\int_0^n\phi_{j,p}(t)\,\mathrm{d}\xi_t\,,$$

where $\phi_{j,p}\;=\;\sum_{k=1}^{np}\phi_j(t_k)\,\mathbf{1}_{(t_{k-1},t_k]}(t)$. In view of the condition $\mathbf{C}_2^*$), this implies that

$$\begin{aligned}\mathbf{E}_S\,(\widehat{\alpha}_{j,p}-\alpha_{j,p})^2 =\;&\frac{1}{p^2}\left(\sum_{k=1}^{p}\phi_j(t_k)\,h_k(S)\right)^2\\&+\frac{1}{n^2}\,\mathbf{E}_S\,\left(\int_0^n\phi_{j,p}(t)\,\mathrm{d}\xi_t\right)^2\le\;H_p(S)+\frac{\lambda^*}{n}\,.\end{aligned}$$

Corollary 3. *Under the conditions* $\mathbf{C}_1^*$)–$\mathbf{C}_2^*$) *the LSE procedure* $\widehat{S}_{\widehat{m}_p,p}$ *satisfies the inequality*

$$\mathbf{E}_S\,\|\widehat{S}_{\widehat{m}_p,p}-S\|_p^2 \le \inf_{m\in\mathcal{M}_p}\widehat{b}_{m,p}(S)+8H_p(S)+\frac{2\tau_0\lambda^*}{n}\,, \tag{45}$$

where

$$\widehat{b}_{m,p}(S)=7\|S_{m,p}-S\|_p^2+7d_mH_p(S)+\lambda^*(7+32z_*\,l_m)\frac{d_m}{n}\,.$$

Now we consider the estimation problem for the model (1)–(2) on the basis of discrete data in the asymptotic setting. First, for any $\beta \geq 1$, we set

$$\mathcal{R}_{n,p}(\widetilde{S}_n, \beta) = \sup_{S \in \Theta_{\beta,r}} \sup_{Q \in \mathcal{P}_\kappa} \mathbf{E}_{S,Q} \|\omega_n (\widetilde{S}_n - S)\|_p^2 \,, \tag{46}$$

where the set $\Theta_{\beta,r}$ is defined by (27) with the use of the trigonometric basis (31), $\omega_n = \omega_n(\beta) = n^{\frac{\beta}{2\beta+1}}$ and the set $\mathcal{P}_\kappa$ is defined in (29). As in Section 4, in order to minimize this risk, we apply the least square model selection procedure $\widehat{S}_{\widehat{m}_p,p}$, constructed on the basis of the trigonometric system (31) with the ordered selection, that is, $\mathcal{M}_p = \{m_1, \ldots, m_p\}$ with $m_j = \{1, \ldots, j\}$. In this case $d_{m_j} = j$ and $l_{m_j} = 1$ for $1 \leq j \leq p$.

It is shown in Appendix 6.7, that if $p \geq n^{1/2}$, then for any $\varepsilon > 0$

$$\overline{\lim}_{n\to\infty} \sup_{\beta \geq 1+\varepsilon} \mathcal{R}_{n,p}(\widehat{S}_{\widehat{m}_p,p}, \beta) \, < \, \infty \tag{47}$$

and if $p \geq n^{1/2}$, then for any $\beta \geq 1$

$$\underline{\lim}_{n\to\infty} \inf_{\widetilde{S}_n} \mathcal{R}_{n,p}(\widetilde{S}_n, \beta) \, > 0 \,. \tag{48}$$

It means that the adaptive estimator $\widehat{S}_{\widehat{m}_p,p}$ with the $p \geq n^{1/2}$ (in particular, one can take $p = 2[n^{1/2}] + 1$) is optimal in the sense of the risk (46).

6 Appendix

6.1 Properties of processes (2)–(3)

We start with the result for process (2) which shows that both conditions $\mathbf{C}_1$)–$\mathbf{C}_2$) and $\mathbf{C}_1^*$)–$\mathbf{C}_2^*$) are satisfied.

Lemma 2. *Let (ξ_t) be defined by (2) with $\theta \leq 0$, $f = (f_j)_{j\geq 1}$ be a family of linearly independent cadlag 1-periodic functions on $\mathbb{R}$, $I_t(f) = \int_0^t f(u) \mathrm{d}\xi_u$. Then the matrix*

$$V_{k,n}(f) = (v_{i,j}(f))_{1\leq i,j\leq k} \tag{49}$$

with elements $v_{i,j}(f) = \mathbf{E}\, I_n(f_i) I_n(f_j)$ is positive definite for each $k \geq 1$, $n \geq 1$ and $\theta \leq 0$. Moreover, if $(f_j)_{j\geq 1}$ is orthonormal, then for any $\theta \leq 0$

$$\sup_{k\geq 1} \sup_{n\geq 1} \sup_{|z|=1} \frac{1}{n} z' V_{k,n}(f) z \leq 2 \,. \tag{50}$$

Proof. Assume that for some $n \geq 1$, $k \geq 1$ and $z = (z_1, \dots, z_k)' \in \mathbb{R}^k$ $z' V_{k,n}(f) z = 0$. Since

$$z' V_{k,n}(f) z = \mathbf{E}\, I_n^2(g)\,,$$

where $g(t) = \sum_{j=1}^k z_j f_j(t)$, one gets $I_n(g) = \int_0^t g(t) \mathrm{d}\xi_t = 0$ a.s. Taking into account that the distribution of $I_n(g)$ for model (2) is equivalent to that of the random variable $\int_0^n g(t) \mathrm{d}w_t$ this implies that $g(t) = 0$ for all $t \in [0, n]$. Thus $z_1 = \ldots = z_k = 0$ and to we come the first assertion. Let us check (50). By applying Ito's formula one obtains

$$\mathbf{E}\, I_n^2(g) = 2\theta \int_0^n g(t)\, \mathbf{E} I_t(g)\, \xi_t \mathrm{d}t + \int_0^n g^2(t) \mathrm{d}t\,,$$

where

$$\mathbf{E} I_t(g)\, \xi_t = \frac{1}{2} \int_0^t g(u) e^{\theta(t-u)} \mathrm{d}u\,.$$

Therefore,

$$\mathbf{E} I_n^2(g) = \theta \int_0^n e^{\theta v} \int_v^n g(t) g(t-v) \mathrm{d}t\, \mathrm{d}v + \int_0^n g^2(t) \mathrm{d}t\,. \tag{51}$$

From here, it follows that for any $\theta \leq 0$

$$\begin{aligned} z' V_{k,n}(f) z &\leq \int_0^n g^2(t) \mathrm{d}t \left(1 + \theta \int_0^\infty e^{\theta v} \mathrm{d}v \right) \\ &\leq 2n \int_0^1 g^2(t) \mathrm{d}t = 2n \sum_{j=1}^k z_j^2 = 2n\,. \end{aligned}$$

This completes the proof of Lemma 2. $\square$

Lemma 3. *Let (ξ_t) be defined by (3) with $\theta \in \mathcal{A}$, $f = (f_j)_{j\geq 1}$ be a family of linearly independent cadlag 1-periodic functions on $\mathbb{R}$. Then the matrix (49) is positive definite for each $k \geq 1$, $n \geq 1$ and $\theta \in \mathcal{A}$. Moreover, if $(f_j)_{1\leq j\leq k}$ is orthonormal, then for any $0 < \delta < 1$ and $\theta \in K_\delta$*

$$\sup_{k\geq 1} \sup_{n\geq 1} \sup_{|z|=1} \frac{1}{n} z' V_{k,n}(f) z \leq \lambda^*(\delta)\,, \tag{52}$$

where $\lambda^*(\delta)$ is defined in (8).

Proof. Let η_t be process (3) with zero initial values, i.e.

$$\mathrm{d}\eta_t^{(q-1)} = \left(\sum_{j=1}^{q} \theta_j \eta_t^{(q-j)}\right) \mathrm{d}t + \mathrm{d}w_t$$

and $\eta_0 = \ldots = \eta_0^{(q-1)} = 0$. Then ξ_t can be written as

$$\xi_t =< e^{At} Y >_q + \eta_t\,, \tag{53}$$

where $< X >_i$ denotes the ith component of a vector X; A is the matrix defined in (5) and Y is a gaussian vector in $\mathbb{R}^q$ independent of $(\eta_t)_{t\ge 0}$ with zero mean and covariance matrix

$$F = \int_0^\infty e^{Au}\, D_q\, e^{A'u}\, \mathrm{d}u\,, \tag{54}$$

where D_q is $q \times q$ matrix in which all elements exept of the $(1,1)$ element are equal to zero ant the $(1,1)$ element is equal to 1. In view of (53), one has

$$I_n(g) = \zeta + \int_0^n g(t) \mathrm{d}\eta_t\,, \tag{55}$$

where $\zeta =< \int_0^n g(t) A e^{At} \mathrm{d}t Y >_q$. Integration by parts yields

$$\int_0^n g(t)\mathrm{d}\eta_t = \int_0^n G_{q-1}(t) \mathrm{d}\eta_t^{(q-1)}\,,$$

where $G_0(t) = g(t)$ and $G_j(t) = \int_t^n G_{j-1}(u)\mathrm{d}u$ for $1 \le j \le q-1$.

Now assume that for some $n \ge 1$, $k \ge 1$ and $z = (z_1, \ldots, z_k)' \in \mathbb{R}^k$ $z' V_{k,n}(f) z = 0$. Since

$$z' V_{k,n}(f) z = \mathbf{E}\, I_n^2(g) = \mathbf{E}\, \zeta^2 + \mathbf{E}\, \left(\int_0^n g(t) \mathrm{d}\eta_t\right)^2\,,$$

this implies that

$$\mathbf{E}\, \left(\int_0^n g(t) \mathrm{d}\eta_t\right)^2 = \mathbf{E}\, \left(\int_0^n G_{q-1}(t) \mathrm{d}\eta_t^{(q-1)}\right)^2 = 0\,.$$

Taking into account that the distribution of the process $(\eta_t^{(q-1)})_{0\le t\le n}$ in $\mathbf{C}[0,n]$ is equivalent to Wiener measure we have

$$\int_0^n G_{q-1}(t) \mathrm{d}w_t = 0 \quad \text{a.s.}$$

and therefore $G_{q-1}(t) = 0$ for all $0 \le t \le n$ and hence $g(\cdot) = 0$ and we obtain $z_1 = \ldots = z_k = 0$. This leads to the first assertion.

Let us show (52). By direct calculations we find

$$\mathbf{E}\, I_n^2(g) = 2\int_0^n < Ae^{Au}FA' >_{q,q} \left(\int_0^{n-u} g(u+s)\, g(s)\mathrm{d}s\right) \mathrm{d}u\,.$$

where $< A >_{i,j}$ denotes the (i,j)-th element of matrix A. By applying the Bunyakovskii-Caushy-Schwartz inequality one gets

$$\mathbf{E} I_n^2(g) \le 2\int_0^n g^2(s)\mathrm{d}s \int_0^n |< Ae^{Au}FA' >_{q,q}|\mathrm{d}u\,.$$

Since

$$\int_0^n g^2(s)\mathrm{d}s = n\int_0^1 g^2(s)\mathrm{d}s = n\sum_{j=1}^k z_j^2 \int_0^1 f_j^2(s)\mathrm{d}s = n\,,$$

we obtain the estimate

$$\frac{1}{n}\, z' V_{k,n}(f) z \, \le 2\int_0^\infty |< Ae^{Au}FA' >_{q,q}|\mathrm{d}u \, \le 2|A|^2|F|J(A)\,,$$

where $J(A) = \int_0^\infty |e^{Au}|\mathrm{d}u$. In order to come to (52) it remains to use the following inequality for matrix exponents of order $q \ge 2$ (see, for example, in Kabanov and Pergamenshchikov (2003) on p. 228)

$$|e^{tB}| \le e^{t\Lambda}\left(1 + 2|B|\sum_{j=1}^{q-1}\frac{1}{j!}(2t|B|)^j\right)\,,$$

where $\Lambda = \max_{1\le j\le q} \mathrm{Re}\,\lambda_j$, λ_j are eigenvalues of the matrix B.

Indeed, from (54) for any $A \in K_\delta$ we find that

$$|F| \le F^*(\delta) \quad \text{and} \quad J(A) \le J^*(\delta)\,,$$

where the functions $F^*(\delta)$ and $J^*(\delta)$ are defined in (8). Hence Lemma 3. $\square$

6.2 Mean forecast inequality

Lemma 4. *(Galtchouk and Pergamenshchikov (2005))*

Let α and ξ be two positive random variables. Let β be a positive real number and $\{\Gamma_x, x \ge 0\}$ be a family of events such that, for any x,

$$\mathbf{P}(\xi > \alpha + \beta x\,,\, \Gamma_x) = 0.$$

Assume also that there exists some positive integrable on $\mathbb{R}_+$ function $M(x)$ dominating $\mathbf{P}(\Gamma_x^c)$. Then

$$\mathbf{E}\,\xi\,\leq\,\mathbf{E}\,\alpha\,+\,\beta M^*\,,$$

where $M^*\,=\,\int_0^\infty\,M(x)\mathrm{d}x$.

Proof. We set $\eta = (\xi - \alpha)_+$. Thus $\mathbf{E}\xi \leq \mathbf{E}\alpha + \mathbf{E}\eta$. Moreover,

$$\begin{aligned}\mathbf{E}\eta &= \int_0^\infty \mathbf{P}(\eta > z)\mathrm{d}z = \int_0^\infty \mathbf{P}(\xi > \alpha + z)\mathrm{d}z \\ &= \beta\int_0^\infty \mathbf{P}(\xi > \alpha + \beta x)\mathrm{d}x \leq \beta\int_0^\infty \mathbf{P}(\Gamma_x^c)\mathrm{d}x \leq \beta M^*\,.\end{aligned}$$

□

6.3 Proof of Theorem 1

By making use of (1) and (10) we obtain

$$\|z - S\|^2 = \gamma_n(z)\,+\,2\,F_n(z)\,+\,\|S\|^2\,, \tag{56}$$

where $F_n(z)\,=\,n^{-1}\,\int_0^n\,z(t)\,\mathrm{d}\xi_t$. Further, from the definition (14), it follows that for each $m \in \mathcal{M}$

$$\gamma_n(\widetilde{S}_{\widetilde{m}})\,\leq\,\gamma_n(\widetilde{S}_m)\,+\,P_n(m) - P_n(\widetilde{m})\,.$$

Thus

$$\|\widetilde{S}_{\widetilde{m}}\,-\,S\|^2 \leq \|\widetilde{S}_m - S\|^2 + 2F_n(\widetilde{z}) + \varpi_n(m,\widetilde{m})\,, \tag{57}$$

where $\widetilde{z} = \widetilde{S}_{\widetilde{m}} - \widetilde{S}_m$ and $\varpi_n(m,\widetilde{m}) = P_n(m) - P_n(\widetilde{m})$. Now for each $x > 0$, $0 < \mu < 1/2\lambda^*$ and set $\iota \in \mathcal{M}$ we introduce the following gaussian function on $\mathcal{D}_\iota + \mathcal{D}_m$

$$U_{x,\iota}(z,\mu) = \frac{2F_n(z)}{\|z\|^2 + \varrho^2_{n,\iota}(x,\mu)}\,,\qquad z \in \mathcal{D}_\iota + \mathcal{D}_m\,, \tag{58}$$

where

$$\varrho_{n,\iota}(x,\mu)\,=\,4\sqrt{\frac{c(\mu)\,N + d_\iota l_\iota + x}{n\mu}}\quad\text{with}\quad c(\mu) = -\frac{1}{2}\,\ln(1 - 2\lambda^*\mu)$$

and $N = \dim(\mathcal{D}_\iota + \mathcal{D}_m)$.

Moreover, let functions $\phi_{i_1},\ldots,\phi_{i_N}$ be the subset of $(\phi_j)_{j\ge 1}$ which is a basis in $\mathcal{D}_\iota+\mathcal{D}_m$. It should be noted that $N\le d_\iota+d_m$. Then one can write a normalized vector $\overline{z}=z/\|z\|$ for $z\ne 0$ as

$$\overline{z}=\sum_{j=1}^N a_j\phi_{i_j}\quad\text{with}\quad\sum_{j=1}^N a_j^2=1\,.$$

Therefore

$$U_{x,\iota}(z,\mu)=\frac{2\,n^{-1/2}\,\|z\|}{\|z\|^2+\varrho^2_{n,\iota}(x,\mu)}\sum_{j=1}^N a_j\,\zeta_{i_j}\quad\text{with}\quad\zeta_j\,=\,\frac{1}{\sqrt{n}}\int_0^n\phi_j\,\mathrm{d}\xi_t\,.$$

By applying the Bunyakovskii-Cauchy-Schvartz inequality one gets

$$|U_{x,\iota}(z,\mu)|\,\le\,\frac{1}{\sqrt{n}\,\varrho_{n,\iota}(x,\mu)}\,\eta_\iota\,,\qquad(59)$$

where $\eta_\iota\,=\,\sqrt{\sum_{j=1}^N\zeta^2_{i_j}}$. Now note that by the condition $\mathbf{C}_1$) the vector $(\zeta_{i_1},\ldots,\zeta_{i_N})$ is gaussian with zero mean and a non-generate covariance matrix B_ι. Therefore

$$\mathbf{E}\,e^{\mu\sum_{j=1}^N\zeta^2_{i_j}}\,=\,\frac{1}{(2\pi)^{N/2}\sqrt{\det B_\iota}}\int_{\mathbf{R}^N}e^{-\frac{1}{2}x'T_\iota^{-1}x}\,\mathrm{d}x\,,$$

where $T_\iota=(B_\iota^{-1}-2\mu I_N)^{-1}$ and I_N is the identity matrix of order N. One can easily verify that

$$\mathbf{E}\,e^{\mu\sum_{j=1}^N\zeta^2_{i_j}}\,=\,\sqrt{\frac{\det T_\iota}{\det B_\iota}}=\frac{1}{\sqrt{\det(I_N-2\mu\,B_\iota)}}\,.$$

Thus, in view of the inequality

$$\det(I_N-2\mu B_\iota)\ge(1-2\mu\,\lambda_{max}(B_\iota))^N$$

and the condition $\mathbf{C}_2$), we obtain

$$\mathbf{E}\,e^{\mu\sum_{j=1}^N\zeta^2_{i_j}}\,\le(1-2\mu\lambda^*)^{-N/2}=e^{c(\mu)N}\,,$$

where $c(\mu)$ is defined in (58).

Now, by the Chebyshev inequality, for any $b>0$ and $0<\mu<1/2\lambda^*$, we obtain that

$$\mathbf{P}(\,\eta_\iota>b\,)\,\le e^{c(\mu)N-\mu b^2}\,.\qquad(60)$$

Choosing in this inequality

$$b = b_*(x,\iota) = \frac{1}{4}\sqrt{n}\,\varrho_{n,\iota}(x,\mu) \, = \sqrt{\frac{c(\mu)\,N + d_\iota l_\iota + x}{\mu}}$$

yields

$$\mathbf{P}(\eta_\iota > b_*(x,\iota)) \leq \, e^{-x-d_\iota l_\iota}\,.$$

Now let $\Gamma_x = \{\sup_{\iota\in\mathcal{M}} \eta_\iota / b_*(x,\iota) \leq 1\}$. It is easy to see that

$$\mathbf{P}(\Gamma^c(x)) \leq \sum_{\iota\in\mathcal{M}} \mathbf{P}(\eta_\iota > b_*(x,\iota)) \leq \sum_{\iota\in\mathcal{M}} e^{-x-d_\iota l_\iota} \, = l^* e^{-x}\,.$$

Thus, we obtain the following upper bound on the set Γ_x

$$\sup_{\iota\in\mathcal{M}} \sup_{z\in\mathcal{D}_\iota+\mathcal{D}_m} |U_{x,\iota}(z,\mu)| \leq 1/4\,,$$

which implies

$$2\,F_n(\widetilde{z}) = U_{x,\widetilde{m}}(\widetilde{z},\mu)(\|\widetilde{z}\|^2 \, + \, \varrho^2_{n,\widetilde{m}}(x)) \leq \frac{1}{2}\|\widetilde{S}_{\widetilde{m}} - S\|^2 \, + \, \frac{1}{2}\|\widetilde{S}_m - S\|^2$$
$$+ \frac{4}{n\mu}(c(\mu) d_m\, l_m \, + \, (c(\mu) \, + \, 1)\, d_{\widetilde{m}} l_{\widetilde{m}}) + \frac{4}{n\mu}\,x\,.$$

By making use of this inequality in (57) we obtain, on the set Γ_x, that

$$\begin{aligned}\|\widetilde{S}_{\widetilde{m}} - S\|^2 &\leq \frac{1}{2}\,\|\widetilde{S}_{\widetilde{m}} - S\|^2 + \frac{3}{2}\|\widetilde{S}_m - S\|^2 + \varpi_n(m,\widetilde{m}) \\ &\quad + \frac{4}{n\mu}(c(\mu) d_m\, l_m + (c(\mu)+1)\, d_{\widetilde{m}} l_{\widetilde{m}}) + \frac{4}{n\mu}\,x \\ &= \frac{1}{2}\,\|\widetilde{S}_{\widetilde{m}} - S\|^2 \, + \, \frac{3}{2}\|\widetilde{S}_m - S\|^2 + \Omega(\rho,\mu) + \frac{4}{n\mu}\,x\,,\end{aligned}$$

i.e.

$$\|\widetilde{S}_{\widetilde{m}} - S\|^2 \leq 3\|\widetilde{S}_m - S\|^2 + 2\Omega(\rho,\mu) + \frac{8}{n\mu}\,x\,,$$

where

$$\Omega(\rho,\mu) = \left(\rho + \frac{4c(\mu)}{\mu}\right)\frac{d_m\, l_m}{n} + \left(\frac{4c(\mu)+4}{\mu} - \rho\right)\frac{d_{\widetilde{m}}\, l_{\widetilde{m}}}{n}\,.$$

It is clear that to obtain a nonrandom minimal upper bound for this term we have to resolve the following optimazation problem

$$\rho + \frac{4c(\mu)}{\mu} \to \min \quad \text{subject to} \quad \frac{4c(\mu)+4}{\mu} - \rho \leq 0\,. \tag{61}$$

One can check directly that the solution of this problem is $\mu = (z_*-1)/2\lambda^* z_*$ and the optimal value for ρ is given in (13). Thus, by choosing these parameters we have on the set Γ_x

$$\|\widetilde{S}_{\widetilde{m}} - S\|^2 \le 3\|\widetilde{S}_m - S\|^2 + 16\lambda^* z_* \frac{l_m d_m}{n} + \frac{16 z_* \lambda^*}{n(z_*-1)}\, x\,.$$

Applying now Lemma 4 with $\xi = \|\widetilde{S}_{\widetilde{m}} - S\|^2$,

$$\alpha = 3\|\widetilde{S}_m - S\|^2 + 16\lambda^* z_* l_m d_m/n\,, \quad \beta = \frac{16 z_* \lambda^*}{n(z_*-1)}$$

and $M(x) = l^* e^{-x}$ we obtain the inequality (16). Hence Theorem 1. $\square$

6.4 Some properties of the Fourier coefficients

Lemma 5. *Let S be a function in $\mathbf{C}^k[0,1]$ such that $S^{(j)}(0) = S^{(j)}(1)$ for all $0 \le j \le k$ and, such that, for some contants $L_0 > 0$, $L > 0$ and $0 \le \alpha < 1$*

$$\max_{0\le j\le k-1} \max_{0\le x\le 1} |S^{(j)}(x)| \le L_0 \quad \textit{and} \quad |S^{(k)}(x) - S^{(k)}(y)| \le L\,|x-y|^{\alpha} \tag{62}$$

for all $x\,, y \in [0,1]$. Then the Fourier coefficients $(a_k)_{k\ge 0}$ and $(b_k)_{k\ge 1}$ of the function S, defined as

$$S(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(2\pi\, kx) + b_k \sin(2\pi\, kx))\,,$$

satisfy the following inequality

$$\sup_{n\ge 0} (n+1)^{\beta} \left(\sum_{j=n}^{\infty} (a_j^2 + b_j^2) \right)^{1/2} \le c^*\,(L + L_0)\,, \tag{63}$$

where $\beta = k + \alpha$ (k being an integer and $0 \le \alpha < 1$) and

$$c^* = 1 + 2^{\beta} + \pi^4\, 9^{\beta}\, \frac{\int_0^{\infty} u^{\alpha-3}\, \sin^4(\pi\, u)\, \mathrm{d}u}{8\, \int_0^{1/2} u^{-4}\, \sin^4(\pi\, u)\, \mathrm{d}u}\,.$$

Proof of this Lemma is given in Fourdrinier and pergamenshchikov (2007) Appendix A4.

6.5 Lower bound for the parametric model

We consider in this section the following model

$$\mathrm{d}y_t = S(t,z)\mathrm{d}t + \mathrm{d}w_t\,, \tag{64}$$

where $(w_t)_{t\geq 0}$ is a standard brownian motion; $z \in \mathbb{R}^m$ is unknown vector parameter. Let now π be a prior distribution density on $\mathbb{R}^m$ of the form

$$\pi(x) = \prod_{l=1}^{m} \pi_l(x_l)\,,$$

where π_l is a positive density on the interval $[-\delta_l, \delta_l]$ for some $\delta_l > 0$. This means that the density π has the following support

$$\Gamma = [-\delta_1, \delta_1] \times \ldots \times [-\delta_m, \delta_m]\,.$$

We set

$$\zeta_l(z) = \int_0^n \frac{\partial}{\partial z_l} S(t,z)(\mathrm{d}y_t - S(t,z)\mathrm{d}t)\,. \tag{65}$$

Now we give some version of the van Trees inequality Gill and Levit (1995) for process (64).

Lemma 6. *For each $l \geq 1$, any estimator $\widetilde{z}_l$ based on observations $(y_t)_{0\leq t\leq n}$ satisfies the following inequality*

$$\int_\Gamma \mathbf{E}_{S_z} (\widetilde{z}_l - z_l)^2\, \pi(\mathrm{d}z) \;\geq\; \frac{1}{A_l + B_l}\,, \tag{66}$$

where $\mathbf{E}_{S_z}$ denotes the expectation with respect to the distribution of process (64),

$$A_l = \int_\Gamma \mathbf{E}_{S_z}\, \zeta_l^2(z)\, \pi(z)\mathrm{d}z \quad \textit{and} \quad B_l = \int_{-\delta_l}^{\delta_l} \frac{(\dot{\pi}_l(u))^2}{\pi_l(u)}\,\mathrm{d}u\,.$$

Proof. It will be noted that the density of the distribution of process (64) with respect to the Wiener measure μ_w on $\mathcal{Y} = \mathbf{C}[0,n]$ is defined as

$$f(y,z) = e^{\int_0^n S_z(t)\,\mathrm{d}y_t - \frac{1}{2}\int_0^n S_z^2(t)\,\mathrm{d}t}\,.$$

Therefore, by applying the method from Gill and Levit (1995) we obtain the lower bound (66). Hence Lemma 6. $\square$

6.6 Proof of Theorem 5

To prove this theorem we adapt the proof of Theorem 1 for this case. In this case equality (56) becomes

$$\|z - S\|_p^2 = \gamma_{n,p}(z) \,+\, 2\,F_{n,p}(z) \,+\, 2\,G_p(z,S) \,+\, \|S\|_p^2\,,$$

with

$$F_{n,p}(z) \;=\; \frac{1}{n}\sum_{k=1}^{np} z(t_k)\,\Delta\xi_{t_k}\,,\quad G_p(z,S) \;=\; \frac{1}{p}\sum_{k=1}^{p} z(t_k)\,h_k(S)\,,$$

where the sequence $h_k(S)$ is defined in (44). Similarly to the proof of Theorem 1, one can show that

$$\begin{aligned}\mathbf{E}_S\|\widetilde{S}_{\widetilde{m}_p,p} - S\|_p^2 \;\le\; & 3\mathbf{E}_S\|\widetilde{S}_{m,p} - S\|_p^2 + 16\lambda^* z_* \frac{d_m l_m}{n} \\ & + 4\mathbf{E}_S|G_p(\widetilde{z}_p,S)| + \frac{\lambda^*\tau_0}{n}\,,\end{aligned} \tag{67}$$

where $\widetilde{z}_p = \widetilde{S}_{\widetilde{m}_p,p} - \widetilde{S}_{m,p}$. Now we note that for any $\nu > 0$

$$\begin{aligned}2G_p(\widetilde{z},S) \;&\le \nu\|\widetilde{z}\|_p^2 + \nu^{-1}\,H_p(S) \\ &\le 2\nu\|\widetilde{S}_{\widetilde{m}_p,p} - S\|_p^2 + 2\nu\|\widetilde{S}_{m,p} - S\|_p^2 + \nu^{-1}\,H_p(S)\,.\end{aligned}$$

Therefore, taking into account the last inequality in (67), we obtain the following upper bound

$$\begin{aligned}\mathbf{E}_S\|\widetilde{S}_{\widetilde{m}_p,p} - S\|_p^2 \;\le\; & \frac{3+4\nu}{1-4\nu}\,\|\widetilde{S}_{m,p} - S\|_p^2 + \frac{16\lambda^* z_*}{1-4\nu}\,\frac{d_m l_m}{n} \\ & + \frac{\lambda^*\tau_0}{(1-4\nu)n} + \frac{2}{\nu(1-4\nu)}\,H_p(S)\,.\end{aligned}$$

By minimzing the last term with respect to ν (i.e. maximazing $\nu(1-4\nu)$) we find that $\nu = 1/8$. Thus the last inequality implies the upper bound (44). Hence, Theorem 5. $\square$

6.7 Proof of (47)

Consider first the principal term in (45). Let $(s_j)_{j\ge1}$ be the Fourier coefficients for S in $\mathcal{L}_2[0,1]$ used in $\Theta_{\beta,r}$. By setting $\Delta_j(t) = S - \sum_{i=1}^{j} s_i\phi_i(t)$, one can estimate $\|S_{m_j,p} - S\|_p^2$ as

$$\|S_{m_j,p} - S\|_p^2 \;=\; \inf_{a_1,\ldots,a_j}\,\|S - \sum_{i=1}^{j} a_j\,\phi_i\|_p^2 \;\le\; \|\Delta_j\|_p^2\,.$$

By the definition of $\varsigma_j(S)$ in (28) we obtain that

$$\|\Delta_j\|_p^2 \le 2\int_0^1 \Delta_j^2(t)\,\mathrm{d}t + 2\sum_{k=1}^p \int_{t_{k-1}}^{t_k} (\Delta_j(t_k) - \Delta_j(t))^2\,\mathrm{d}t$$
$$= 2\,\varsigma_{j+1}(S) + 2\sum_{k=1}^p \int_{t_{k-1}}^{t_k} \left(\int_t^{t_k} \dot{\Delta}_j(u)\,\mathrm{d}u\right)^2 \mathrm{d}t\,. \tag{68}$$

Moreover, the Bunyakovskii-Cauchy-Schwartz inequality implies that

$$\|\Delta_j\|_p^2 \le 2\,\varsigma_{j+1}(S) + \frac{2}{p^2}\,\|\dot{\Delta}_j\|^2\,.$$

Notice now that for the trigonometric basis (31) and for the functions S from $\Theta_{\beta,r}$ with $\beta > 1$ we obtain that for any $j \ge 0$

$$\|\dot{\Delta}_j\|^2 = \sum_{i=j+1}^{\infty} s_i^2\,\|\dot{\phi}_i\|^2 \le \pi^2 \sum_{i=j+1}^{\infty} s_i^2 i^2$$
$$\le \pi^2 (j+1)^2\,\varsigma_{j+1}(S) + 2\pi^2 \sum_{i>j} (i+1)\,\varsigma_{i+1}(S)$$
$$\le r^2\,\pi^2\,(j+1)^{-2(\beta-1)}\,\frac{\beta}{\beta-1}\,. \tag{69}$$

Therefore for $1 \le j \le p$

$$\sup_{S\in\Theta_{\beta,r}} \|S_{m_j,p} - S\|_p^2 \le 2\,\frac{r^2}{j^{2\beta}}\left(1 + \frac{\pi^2 j^2}{p^2}\,\frac{\beta}{\beta-1}\right)\,. \tag{70}$$

Moreover, taking into account that $H_p(S) \le p^{-2}\|\dot{S}\|^2$, through (69) with $j=0$ we get that for $p \ge n^{1/2}$

$$\sup_{S\in\Theta_{\beta,r}} H_p(S) \le \frac{\beta\pi^2\,r^2}{p^2(\beta-1)} \le \frac{\beta\pi^2\,r^2}{n(\beta-1)}\,.$$

Thus (45) implies that for any $\varepsilon > 0$ there exists some constant $C^* = C^*(r,\epsilon) > 0$ such that for any $\beta \ge 1+\epsilon$, $p \ge n^{1/2}$ and for $1 \le j \le p$

$$\mathcal{R}_{n,p}(\widehat{S}_{\widehat{m}_p,p},\beta) \le C^*\left(\left(j^{-2\beta} + \frac{j}{n}\right) n^{\frac{2\beta}{2\beta+1}} + n^{-\frac{1}{2\beta+1}}\right)\,.$$

This bound with $j = j_* = [n^{\frac{1}{2\beta+1}}] + 1$ implies immediately inequality (47).
$\square$

6.8 Proof of (48)

Notice now that for any estimator $\widetilde{S}_n$ by putting $T_p(\widetilde{S}_n)(t) = \sum_{j=1}^{p} \widetilde{S}_n(t_j)\mathbf{1}_{(t_{j-1},t_j]}$ we can represent the accuracy of this estimator as

$$\|\widetilde{S}_n - S\|_p^2 = \sum_{j=1}^{p} \int_{t_{j-1}}^{t_j} (T_p(\widetilde{S}_n)(t) \, - S(t_k))^2 \, \mathrm{d}t \, .$$

Therefore, for any $0 < \epsilon < 1$ we can estimate with below this accuracy as

$$\|\widetilde{S}_n - S\|_p^2 \geq (1-\epsilon)\|T_p(\widetilde{S}_n) - S\|^2 - (\epsilon^{-1} - 1) \sum_{j=1}^{p} \int_{t_{j-1}}^{t_j} (S(t) - S(t_j))^2 \mathrm{d}t \, .$$

Moreover, similarly to (68)–(70) we obtain that

$$\sum_{j=1}^{p} \int_{t_{j-1}}^{t_j} (S(t) - S(t_j))^2 \mathrm{d}t \, \leq \, p^{-2} \, \|\dot{S}\|^2 \, \leq \, r\pi^2 n^{-1} \, .$$

Therefore

$$\mathcal{R}_{n,p}(\widehat{S}_n, \beta) \, \geq \, (1-\epsilon) \inf_{T_n} \mathcal{R}_n(T_n, \beta) \, - \, (\epsilon^{-1} - 1) \, r\pi^2 \, n^{-\frac{1}{2\beta+1}} \, ,$$

where the risk $\mathcal{R}_n(T_n, \beta)$ is defined by (29) for some estimator T_n. Now Theorem 4 directly implies (48). $\square$

References

[1] Barron, A., Birgé, L., Massart, P. (1999) Risk bounds for model selection via penalization. *Probability Theory and Related Fields*, 113, 301-413

[2] Baraud, Y. (2000) Model selection for regression on a fixed design. *Probability Theory and Related Fields*, 117, 467-493

[3] Birge, L., Massart, P. (2001) Gaussian model selection. *Journal of the European Mathematical Society*, 3, 203-268

[4] Chenstov, N.N. (1962) On estimation of unknown density by observations. *Reports of Academy of Sciences of USSR*, 147(1), 45-48

[5] Efroimovich, S.Yu. (1999) *Nonparametric curve estimation. Methods, theory and applications.* Springer-Verlag, Berlin, New York

[6] Fourdrinier, D., Pergamenshchikov, S. (2007) Improved model selection method for a regression function with dependent noise. *Annals of the Institute of Statistical Mathematics*, 59, 435-464

[7] Galtchouk, L., Pergamenshchikov, S. (2005) Efficient adaptive nonparametric estimation in heteroscedastic regression model. *Preprint, IRMA, Université de Strasbourg Louis Pasteur*, 020 (available at *http://www.univ-rouen.fr/LMRS/Persopage/Pergamenchtchikov*)

[8] Galtchouk, L., Pergamenshchikov, S. (2006) Asymptotically efficient estimates for nonparametric regression models. - *Statistics and Probability Letters*, 76(8), 852-860

[9] Golubev G.(1982) Minimax filtering of functions. *Problems of Transimission Information*, 4, 67-75

[10] Golfed, S.M. and Quandt, R.E. (1972) *Nonlinear methods in econometrics.* North-Holland Publishing

[11] Gill, R.D. and Levit, B.Y. (1995) Applications of the van Trees inequality: a Bayesian Cramér–Rao bound.*Bernoulli*, 1, 59-79

[12] Ibragimov, I.A. and Hasminskii, R.z. (1981) *Statistical Estimation: Asymptotic Theory.* Springer, New York

[13] Kabanov, Yu.M. and Pergamenshchikov, S.M. (2003) *Two-Scale Stochastic Systems. Asymptotic Analysis and Control.* Springer, New York

[14] Konev, V.V. Pergamensgchikov, S.M. (2003) Sequential estimation of the parameters in a trigonometric regression model with the gaussian coloured noise. *Statistical Inference for Stochastic Processes*, 6, 215-235

[15] Liptser, R. Sh., Shyraev, A.N. (1977) *Statistics of Random Processes.I. General theory.* Springer-Verlag, New York-Heidelberg

[16] Nemirovskii, A. (2000) Topics in non-parametric statistics. *Lecture Notes in Mathematics*, 1738, 85-277

[17] Pinsker, M.S.(1981) Optimal filtration of square integrable signals in gaussian white noise. *Problems of Transimission Information*, 17, 120–133